

A Kidnapping Detection Using Human Pose Estimation in Intelligent Video Surveillance Systems

Ju Hyun Park*, KwangHo Song**, Yoo-Sung Kim***

Abstract

In this paper, a kidnapping detection scheme in which human pose estimation is used to classify accurately between kidnapping cases and normal ones is proposed. To estimate human poses from input video, human's 10 joint information is extracted by OpenPose library. In addition to the features which are used in the previous study to represent the size change rates and the regularities of human activities, the human pose estimation features which are computed from the location of detected human's joints are used as the features to distinguish kidnapping situations from the normal accompanying ones. A frame-based kidnapping detection scheme is generated according to the selection of J48 decision tree model from the comparison of several representative classification models. When a video has more frames of kidnapping situation than the threshold ratio after two people meet in the video, the proposed scheme detects and notifies the occurrence of kidnapping event. To check the feasibility of the proposed scheme, the detection accuracy of our newly proposed scheme is compared with that of the previous scheme. According to the experiment results, the proposed scheme could detect kidnapping situations more 4.73% correctly than the previous scheme.

▶Keyword: Intelligent Video Surveillance Systems, Kidnapping Detection, Human Pose Estimation, Machine Learning, Classification Schemes

1. Introduction

경찰청 통계자료에 따르면, 납치, 폭행 등을 포함한 범죄의 발생 빈도가 최근 5년 동안 지속적으로 증가하고 있는 것으로 보고되고 있다[1]. 이러한 범죄의 관련 증거를 확보하고 더 나아가 범죄 발생을 미연에 방지하기 위해 지능형 영상 감시 시스템(intelligent video surveillance system)에 대한 연구와 활용이 증가되고 있는 추세이다[2,3].

지능형 영상 감시 시스템에서 납치 상황의 발생을 인식하기 위해서 과거 납치 상황 인식을 위한 과거의 연구에서는 납치 상황에서 납치범(kidnapper)과 피해자(victim) 모두의 행위가

일반 상황에서 사람들의 행위보다 크고 불규칙적이라는 특성에 따라 비디오 영상으로부터 사람들의 위치 및 크기 정보를 추출하여 기계 학습(machine learning)의 분류 방법(classification scheme)으로 납치 상황과 일반 동행 상황을 구분하는 납치 상황 인식기를 제안하였다[4]. 그러나 비디오 영상으로부터 사람들의 위치 및 크기 정보를 추출하여 이를 기반으로 납치 상황과 일반 동행 상황을 정확하게 구분하는 데에는 한계가 있을 수 있다. 왜냐하면, 납치 상황과 일반 동행 상황을 구분하기 위한 사람들의 행위 크기와 규칙성 정도의 일반적인 분리 기준을

• First Author: Ju Hyun Park, Corresponding Author: Yoo-Sung Kim

*Ju Hyun Park (bluesping3@gmail.com), Dept. of Information and Communication Engineering, Inha University

**KwangHo Song (crossofjc@gmail.com), Dept. of Information and Communication Engineering, Inha University

***Yoo-Sung Kim (yskim@inha.ac.kr), Dept. of Information and Communication Engineering, Inha University

• Received: 2018. 07. 03, Revised: 2018. 07. 16, Accepted: 2018. 07. 25.

• This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT). (2017-0-00336, Platform Development of Multi-log based Multi-Modal Data Convergence Analysis and Situational Response)

정하는 것이 어렵고, 납치범과 피해자의 행위 크기가 작고 규칙성이 있는 납치 상황도 존재하고 반대로 행위 크기가 크고 규칙성이 없는 일반 동행 행위도 존재할 수 있기 때문이다.

따라서 납치 상황을 정확하게 인식하기 위해서는 행위의 크기와 규칙성뿐만 아니라 사람의 행위 자세를 정확하게 인식하여 이용하는 방법이 필요하다. 사람의 행위 자세를 인식하기 위해 사람의 관절정보를 활용하여 납치를 비롯한 여러 이상 상황들을 인식하는 연구가 존재한다[5]. 또한 이 연구에서는 키넥트(kinect)의 깊이 카메라(depth camera)를 이용하여 관절의 위치 정보를 취득하고 사람들의 관절의 위치변화 패턴으로 납치 상황을 인지하는 이상행동 인식 기법을 제안하였다[5]. 그러나 키넥트는 일반적으로 실내에서 가까운 거리의 상황 인식을 위한 목적으로 사용하기 때문에 실외에서 원거리의 위치에서 발생한 납치 상황을 인식하기 위해서 키넥트를 사용하여 사람들의 관절 정보를 취득하는 것은 어렵다 [6].

따라서 주로 실외에서 사용될 지능형 영상 감시 시스템을 위한 납치 상황 인식기에서는 많은 제약이 따르는 키넥트를 사용할 수 없고 영상으로부터 사람의 관절 정보를 취득할 수 있는 방법이 필요하다. 본 논문에서는 영상 만으로부터 사람의 관절 정보를 정확하게 구하기 위해서 다양한 자세의 영상 이미지와 관절 정보를 추출할 수 있도록 개발된 OpenPose 라이브러리[7,8,9]를 이용하여 사람의 관절을 인식하도록 하고, 과거 납치 상황 인식에서 사용한 영상으로부터 추출한 특징 정보들[4]과 함께 영상으로부터 인식한 사람의 관절 정보로 구성된 특징 정보들을 추가로 이용해서 납치 상황에서의 납치범과 피해자의 행위 자세와 일반 상황에서 사람들의 동행 자세와 구분하는 납치 상황 인식기를 제안하였다.

본 연구에서는 납치 상황 인식기 개발을 위해 필요한 학습 및 테스트 데이터로 사용하기 위해서 YouTube 사이트에서 납치 행위 동영상을 검색하여 참고하였으며 필요한 실험 데이터를 확보하기 위해 검색된 동영상과 유사한 상황의 납치 동영상과 그리고 일반 동행 동영상을 촬영하여 학습 및 테스트를 위한 비디오 데이터로 사용하였다. 납치 동영상과 일반 동행 동영상을 입력 받아서 추출한 과거 납치 상황 인식에서 사용한 영상 특징들[4]에 추가하여 OpenPose 라이브러리를 이용하여 사람의 관절을 인식하고 각 입력 비디오로부터 인식된 사람 관절의 개수, 인식된 관절의 거리 및 각도, 관절의 이동 속도, 폐색(Occlusion) 유·무 등의 특징을 후보 특성으로 정의하고 분별력이 높은 속성만을 선택하여 입력으로 사용하였다. 또한, 기계 학습 지원 도구인 WEKA[11]에서 제공하고 있는 다양한 분류 기법들을 이용하여 각각의 납치 상황 인식기를 설계하고 정확도 분석 실험을 통해 제일 높은 인식 정확도를

갖는 납치 상황 인식기 모델을 제안하였다. 또한, 본 연구에서 제안하는 사람 자세 추정을 추가로 이용하는 납치 상황 인식기의 효용성을 평가하기 위해 기존의 납치 상황 인식기[4]와의 인식 정확도를 비교하는 실험을 통해 본 논문에서 제안하는 납치 상황 인식기가 기존의 납치 상황 인식기보다 정확하게 납치 상황을 인식할 수 있음을 확인하였다.

본 논문의 이후 구성은 다음과 같다. 2장에서는 납치 상황 인식기에 관한 기존 연구에 대해 간략히 설명하고 3장에서는 본 논문에서 제안하는 사람 자세 추정을 이용한 납치 상황 인식기에 대해 자세히 설명한다. 4장에서는 제안된 사람 자세 추정을 이용한 납치 상황 인식기의 성능을 평가하기 위해서 기존의 납치 상황 인식 연구에서 제안되었던 방법[4]과의 인식 정확도 비교 분석에 대해 기술한다. 마지막으로 5장에서는 본 논문의 결론을 맺는다.

II. Related works

본 장에서는 지능형 영상 감시 시스템에서 납치 상황의 발생을 인식하기 위한 연구[4]에 대해 설명하고, 이어서 사람의 관절 정보를 이용한 이상행동 예측 연구[5] 대해 간략히 설명한다.

납치 상황 인식기 개발의 기존 연구를 수행한 납치 상황 인식에서는 납치 상황에서는 납치범과 피해자의 행위가 일반 상황에서 동행하는 사람들의 행위보다 크기가 비교적 크고 불규칙적으로 이뤄지는 움직임에 따라 납치와 동행을 구분하였다[4]. 이를 위해서 Fig. 1과 같이 입력 비디오의 각 프레임으로부터 사람을 인식하여 사람 ROI(Region of Interest)로 표시하고 두 사람이 조우한 프레임으로부터는 두 사람이 함께 인식되어 하나의 그룹 ROI로 표시된다. 입력 비디오에서 두 사람이 처음으로 조우한 프레임 이전에서의 사람 ROI의 크기와 움직임 속도 등과 조우한 이후의 그룹 ROI의 크기 변화 및 속도 변화의 패턴으로 납치를 인식하였다. 납치 상황을 인식하기 위해 고려한 후보 특징은 Table 1에 기술되어 있으며, WEKA의 특징 선택(attribute selection) 기능을 이용하여 이들 중에서 납치 상황 인식에 효과적으로 사용될 수 있는 분별력이 높은 속성으로 X_rate, X_rate+Y_rate, Max_width_rate, Max_height_rate을 선정하였다.

입력 비디오에서 두 사람이 처음으로 조우한 이후의 프레임들에 대해서 납치 상황의 프레임인지 여부를 인식하기 위해 선택된 속성을 이용하여 다양한 기계 학습의 분류 방법으로 프레



Fig. 1. Kidnapping detection process of [4]

임 단위 납치 상황 인식기 모델을 설계하고 인식의 정확성을 평가한 결과에 따르면 베이지안 네트워크(Bayesian network) 기반의 분류 모델이 제일 높은 인식 정확도를 갖는 것으로 판단되었다. 따라서 기존의 납치 상황 인식 연구에서 제안한 납치 상황 인식기는 입력 비디오 영상에서 두 사람이 조우한 이후의 프레임들에 대해서 프레임 단위로 납치 상황 발생 여부를 분류하여, 전체 프레임의 수 대비 납치 상황 발생 프레임의 비율이 일정 문턱치를 넘는 경우, 해당 비디오는 납치 상황이 발생한 것으로 판단하였다[4]. 그러나 이 방법을 이용하여 납치 상황 비디오와 일반 동행 상황 비디오를 정확하게 구분하는 데에는 어려움이 있다. 왜냐하면, 납치 상황과 일반 동행 상황을 구분하기 위한 사람들의 행위 크기와 규칙성 정도의 일반적인 분리 기준을 정하는 것이 어렵고, 납치범과 피해자의 행위 크기가 작고 규칙성이 있는 납치 상황도 존재하고, 반대로 행위 크기가 크고 규칙성이 없는 동행 행위도 존재할 수 있기 때문이다. 따라서 납치 상황을 정확하게 인식하기 위해서는 행위의 크기와 규칙성뿐만 아니라 사람의 행위 자세를 정확하게 인식하여 이용하는 방법이 필요하다.

Table 1. Candidate features for classifying between kidnapping case and normal one

Attribute Name	Description
Width_rate	Change rate of ROI width to the previous frame
Height_rate	Change rate of ROI height to the previous frame
Width_rate + Height_rate	Sum of Width_rate and Height_rate
X_distance	Difference of ROI X-axis value from the previous frame
Y_distance	Difference of ROI Y-axis value from the previous frame
All_distance	Difference of ROI values from the previous frame
X_rate	Change rate of ROI X-axis value to the previous frame
Y_rate	Change rate of ROI Y-axis value to the previous frame
X_rate + Y_rate	Sum of X_rate and Y_rate
Width_sub	Difference of ROI width from the previous frame
Height_sub	Difference of ROI height from the previous frame
Width_sub + Height_sub	Sum of Width_sub and Height_sub
Max_width_rate	(Width of ROI after encountering) / (Sum of maximum widths of two people before encountering)
Max_height_rate	(Height of ROI after encountering) / (Maximum value of two people's height before encountering)

사람의 관절 정보를 이용하여 납치 상황 인식을 제안한 여러 이상 상황들을 수행한 기존 연구에서는 키넥트의 깊이 카메라를 이용하여 사람 관절의 위치 정보를 취득하고 사람들의 관절의 위치와 움직이는 변화 패턴을 기반으로 납치 상황 발생을 인지하는 이상행동 예측을 제안하였다[5]. 또한 납치 상황 등과 같이 이상행동을 인지하기 위해 좀 더 세분화된 팔의 움직임 패턴과 몸의

움직임 패턴을 이용하였다[5]. 예를 들면, 일반적으로 사람이 걷는 경우에는 팔은 허벅지 부근에서 앞·뒤로 약간의 흔들림이 있고 팔꿈치부터 어깨까지의 위치는 거의 움직이지 않는 형태를 나타내며 몸은 하체 움직임이 일정한 반면에, 이상 행동을 하는 경우에는 보통 걸었을 때보다 팔 전체에서 큰 움직임을 보이며 사람 간에 부딪히는 행위가 발생되고, 몸의 움직임은 상대를 향해 기울어지게 되는 사실을 근거로 해서 보통 상황과 이상 상황을 구분하였다. 이러한 상황들을 수행한 기존 연구에서 팔의 움직임 패턴을 인식하기 위해 사용하는 관절은 어깨, 팔꿈치, 허리이며, 그리고 몸의 움직임 패턴을 인식하기 위해서 사용하는 관절은 머리, 어깨, 발목을 이용하였다[5]. 그러나 이러한 상황들을 수행한 기존 연구는 사람의 관절 정보를 얻기 위해 키넥트의 깊이 카메라를 사용하고 있는데 [5] 키넥트는 주로 실내의 근거리를 촬영하는 목적으로 많이 이용되지만 실외 원거리의 위치한 사람의 관절을 정확하게 인식하는 데에는 많은 제약이 있다[6].

영상으로부터 사람의 관절 정보를 정확하게 구하기 위한 사람 자세 추정에 관한 기존 연구가 있다[7,8]. 특히 사람의 움직이는 자세를 추정하여 움직인 관절을 구분하기 위해 수행한 연구에서는 여러 사람의 포즈를 동시에 인식하기 위해 각 사람의 해부학적 중요 위치로서 관절들의 위치를 구하고 관절의 위상 및 연결 관계 등을 이용하여 각 사람의 관절 정보를 실시간으로 구분해 내는 방법을 제안하였으며[8] 이러한 연구 결과를 구현하여 OpenPose 라이브러리로 발표하였다[9]. 따라서 본 연구에서는 납치 상황에서 납치자와 피해자의 정확한 행위를 파악하기 위해서 영상으로부터 OpenPose 라이브러리를 이용하여 각 사람의 관절의 위치 정보를 구하고 관절간의 거리, 이동 속도, 폐색 여부 등으로부터 구한 관절 특징 정보를 기존 연구인 납치 상황 인식에서 사용한 ROI 크기 변화로 유추한 행위 크기 및 규칙성 특징 정보[4]에 추가하여 납치 상황에서의 납치자와 피해자의 행위와 일반상황에서의 동행자들의 행위를 구분하여 납치 상황을 인식하는 방법을 제안한다.

III. A kidnapping detection using human pose estimation

본 장에서는 기존의 납치 상황 인식 연구에서 제안한 납치 상황 인식기[4]의 단점을 보완하기 위해 영상 내 행위자가 만드는 행위의 규모나 변화량뿐만 아니라 행위자의 관절 정보를 추가적으로 고려한 새로운 납치 상황 인식 모델을 제안한다. 이를 위해 OpenPose 라이브러리[9]를 사용해 영상 내 사람의 관절의 위치정보를 추출하고 이를 가공하여 만들어낸 특징들을 추가, 모델 형성에 활용한 새로운 납치 행위 인식 모델을 제안한다.

3.1 Overall procedure

본 연구에서 제안하는 납치 상황 인식 모델은 Fig. 2와 같은

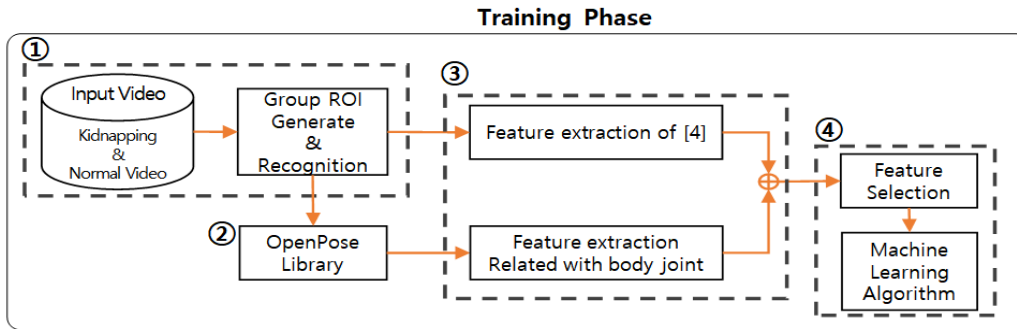


Fig. 2. Overall procedure of training the proposed kidnapping detection scheme

학습과정을 거쳐 만들어진다. 먼저 Fig. 1과 같이 영상에 등장한 사람들의 영역을 찾아 각각의 사람 ROI로 표시하고 동시에 납치 인식에 필요한 특징 값들을 수집하며, 두 사람의 ROI들이 영상에서 조우한 후 이동하는 경우에는 이를 그룹 ROI(Group ROI)로 표시하여 추적하는 단계(Fig. 2의 ①)를 거쳐 필요한 ROI들을 만들어 낸다. 이어서 두 번째 단계에서는 Fig. 2의 ②와 같이 그룹 ROI 영역의 이미지에 OpenPose 라이브러리[9]를 적용하여 ROI 영역 내에 존재하는 행위자들의 신체 관절의 위치 정보를 추출한다. 이어지는 세 번째 단계에서는 기존 연구에서 선정한 4가지 중요 특징(X_rate, X_rate+Y_rate, Max_width_rate, Max_height_rate)[4]과 별도로 앞서 추출한 그룹 ROI내 행위자의 신체 관절의 위치정보를 토대로 다양한 후보 특징들의 계산한다(Fig. 2의 ③). 추출된 후보 특징들 중에서 납치 상황 인식에 크게 기여하는 중요한 최종 특징들을 선별해 내는 작업을 수행한 후, 마지막으로 앞서 선별된 최종 특징들을 바탕으로 다양한 기계학습 알고리즘을 활용해 최적의 납치상황 인식 모델을 학습한다(Fig. 2의 ④).

인식에 중요한 역할을 할 수 있음을 보인 ROI의 변화량 관련 특징으로 2장 관련연구에서 설명한 바와 같이

3.2 Joint detection using OpenPose library

앞서 간략히 설명한 바와 같이 본 연구에서 고려하는 후보 특징 중 하나인 행위자의 신체 자세, 몸과 팔의 움직임 등을 모델에 반영하기 위해서는 행위자의 신체를 구성하는 관절들의 위치나 움직임 정보가 필요하다. 다만 원거리의 영상정보만 활용 가능한 CCTV의 특성상 [5]와 같이 키넥트 등의 센서를 이용하기 어렵기 때문에 본 연구에서는 OpenPose 라이브러리 [9]를 이용하여 영상으로부터 행위자의 신체 관절의 위치정보를 영상 내의 (x, y) 좌표로 취득한다.

OpenPose 라이브러리를 이용하여 취득 가능한 사람의 관절 위치는 총 18개 영역이며, 이들 중에서 본 논문에서는 기존 연구에서 납치 상황 인식에 중요한 역할을 할 수 있다고 제시된 [5] 팔의 움직임과 관련된 관절(Fig. 3의 왼쪽)과 몸의 움직임을 알아 낼 수 있는 관절(Fig. 3의 오른쪽)로 총 10개 관절의 위치 정보를 구하여 이용한다.

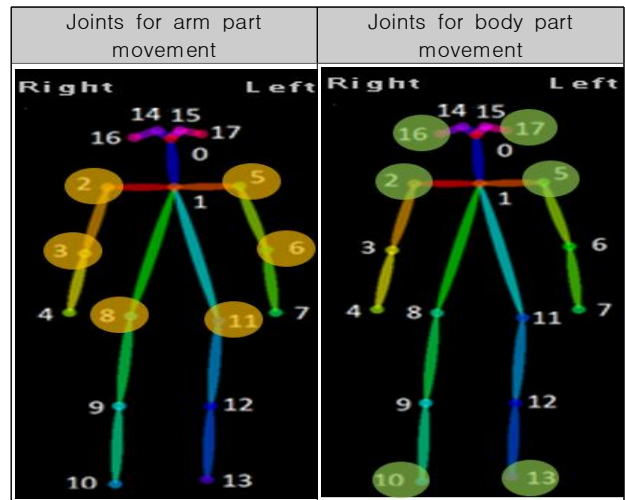


Fig. 3. Selected joints for arm part movement and body part movement

3.3 Feature extraction for kidnapping detection

본 연구에서 고려하는 후보 특징의 종류는 크게 두 가지로 나뉘는데, 첫째는 기존의 납치 상황 인식 연구에서 납치 상황

Table 1의 특징 중에서 중요한 특징으로 선택된 4가지 속성(X_rate, X_rate+Y_rate, Max_width_rate, Max_height_rate)[4]이고 둘째는, 앞서 설명한 바와 같이 여러 이상 상황들 중에서 그 중 납치 상황 인식에 중요한 역할을 할 수 있다고 제시된[5] 신체 관절 10개의 위치정보들로부터 식(1)~(2)를 적용하여 계산한 몸과 팔의 움직임 관련 특징이다.

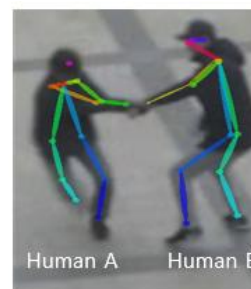


Fig. 4. Example of human parts recognition using OpenPose library

먼저 식 (1)은 Fig. 4와 같이 그룹 ROI에 등장하는 두 행위

자 A, B 사이의 대칭되는 동일 관절 부위 사이의 유클리드 거리(Euclidean Distance)를 계산하는 식이다[10].

$$EuclideanDistance(A_{dir}^{joint}, B_{dir}^{joint}) \quad (1)$$

$$= \sqrt{(x_{(joint,dir)}^A - x_{(joint,dir)}^B)^2 + (y_{(joint,dir)}^A - y_{(joint,dir)}^B)^2}$$

여기서 A_{dir}^{joint} 는 행위자 A의 좌측 또는 우측(dir로 구분)에 해당하는 특정 관절(joint로 표기)을 의미하며 $x_{(joint,dir)}^A$ 와 $y_{(joint,dir)}^A$ 는 그 관절의 x와 y 좌표값을 의미한다. 이 식으로 계산되는 특성은 행위자들 사이의 마주보는 동일 관절들의 거리가 불규칙적으로 자주 변화되는 패턴을 보이는 납치 상황의 특성을 반영하기 위한 것으로서 Table 2에 관절 거리 특성 그룹에 속하는 특징들이다.

Table 2. Candidate features based on Euclidean distance of the corresponding joints of two people in group ROI

Feature Name	From(A)		To(B)	
	Joint	Dir	Joint	Dir
EucDist_LR_Ear	Ear	Left	Ear	Right
EucDist_RL_Ear	Ear	Right	Ear	Left
EucDist_LR_Shoulder	Shoulder	Left	Shoulder	Right
EucDist_RL_Shoulder	Shoulder	Right	Shoulder	Left
EucDist_LR_Elbow	Elbow	Left	Elbow	Right
EucDist_RL_Elbow	Elbow	Right	Elbow	Left
EucDist_LR_Hip	Hip	Left	Hip	Right
EucDist_RL_Hip	Hip	Right	Hip	Left
EucDist_LR_Ankle	Ankle	Left	Ankle	Right
EucDist_RL_Ankle	Ankle	Right	Ankle	Left

식 (2)는 Fig. 4와 같이 그룹 ROI에 등장하는 두 행위자 A, B 사이의 대칭되는 동일 관절 부위 사이의 위치 변화 속도(Velocity)를 계산하는 식이다[10].

$$Velocity(A_{dir}^{joint}, B_{dir}^{joint}) \quad (2)$$

$$= \frac{EuclideanDistance(A_{dir}^{joint}, B_{dir}^{joint})}{fps}$$

여기서 fps(frames per second)는 상황 인식의 대상이 되는 영상의 fps를 이용한다. 이 식으로 계산되는 특성은 행위자들 사이의 마주보는 동일 관절들의 거리의 변화량이 크고 그 방향이 자주 변화한다는 특징을 보이는 납치상황의 특성을 반영하기 위한 것으로서 Table 3에 관절 속도 그룹에 속하는 특징들이다.

이렇게 계산된 신체 관절의 위치정보와 관련된 20가지 특징(Table 2와 Table 3 참조)들을 납치상황 인식을 위한 자세를 추정하기 위한 후보 특징으로 추가한다. 한편, 영상의 촬영 각도에 따라 또는 납치 상황의 경우 그룹 ROI내 행위자들의 격한 움직임에 따라 영상에서 특정 관절이 나타나지 않는 폐색현상(occlusion)이 자주 발생할 수 있으며, 이 경우 해당 시점에 특

정 관절과 관련한 특징 값들의 계산에 오류가 발생 할 수 있다. 식 (3)은 폐색이 발생하여 관절을 인식하지 못한 비율을 표시하는 Occlusion_Rate를 계산하는 식이다.

Table 3. Candidate features based on moving velocity of the corresponding joints of two people in group ROI

Feature Name	From(A)		To(B)	
	Joint	Dir	Joint	Dir
Velocity_LR_Ear	Ear	Left	Ear	Right
Velocity_RL_Ear	Ear	Right	Ear	Left
Velocity_LR_Shoulder	Shoulder	Left	Shoulder	Right
Velocity_RL_Shoulder	Shoulder	Right	Shoulder	Left
Velocity_LR_Elbow	Elbow	Left	Elbow	Right
Velocity_RL_Elbow	Elbow	Right	Elbow	Left
Velocity_LR_Hip	Hip	Left	Hip	Right
Velocity_RL_Hip	Hip	Right	Hip	Left
Velocity_LR_Ankle	Ankle	Left	Ankle	Right
Velocity_RL_Ankle	Ankle	Right	Ankle	Left

$$Occlusion_Rate = \frac{NUMofUndetectedjoint}{MAXNUMofDetectedjoint} \quad (3)$$

여기서 $MAXNUMofDetectedjoint$ 는 관찰대상이 되는 관절들의 최대 탐지 개수로서 앞서 납치 상황 인식에서 중요한 역할을 할 수 있다고 제시된 관절들의 수인 10을 의미하며 $NUMofUndetectedjoint$ 는 해당 시점에 폐색으로 인해 탐지되지 않은 관절의 수를 의미한다.

이어서 Table 2와 Table 3에서 정의된 후보 특징 그리고 Occlusion_Rate를 포함한 총 21가지의 새로운 사람 자세 추정 후보 특징과 납치 상황 인식에 중요하다고 제시한 4가지 행위 크기 및 규칙성 특징[4]을 더한 25개 최종 후보특징을 대상으로 이들 중 납치 상황 인식에 효과적인 특성들을 선별하기 위해 특징 선정(feature selection)을 수행하였다. 그 결과, 총 8가지 특징 X_Rate, X_Rate + Y_rate, Maxwidth_rate, Maxheight_rate, EucDist_LR_Shoulder, EucDist_LR_Elbow, EucDist_LR_Ankle, Occlusion_Rate)이 납치 상황 인식에 유효한 최종 특징으로 선정되었다. 이때 속성 선정을 위한 가치 평가 방법으로는 WEKA[11]에서 제공되는 CfsSubsetEval(Correlation-based Feature Subset Selection)[12]을 이용하였고 탐색 방법으로는 BestFirst[13] 방법을 사용하였다.

선정된 8가지 특징들의 구성을 살펴보면 납치 상황 인식에 활용된 4가지의 특징 (X_Rate, X_Rate + Y_Rate, Maxwidth_rate, Maxheight_rate)[4]이 모두 포함되었고, 더불어 납치상황 인식을 위한 행위자의 움직임 패턴을 관찰하는데 있어 중요한 관절들[5]로 꼽은 어깨, 팔꿈치 그리고 발목의 위치정보에 관한 특징 3가지 EucDist_LR_Shoulder, EucDist_LR_Elbow, EucDist_LR_Ankle 그리고 납치 상황과 같이 격한 움직임을 보이는 영상에서 많이 일어나는 폐색과 관련된 특징 Occlusion_Rate이 선정되었으며, 각 그룹의 후보 특징들이 골고루 선정되었음을 알 수 있다.

앞서 선정된 8가지 특징을 기반으로 영상의 각 프레임이 납치상황인지 아닌지를 분류(classification)하기 위한 모델을 만

들기 위해 기계학습 알고리즘 중 확률을 기반으로 분류하는 Naive Bayes[14], 결정 트리 기반으로 분류하는 J48 decision tree(C4.5)[15], 데이터의 분포를 기반으로 분류하는 SVM(Support Vector Machine)[16] 각각을 학습하여 인식 모델을 생성하고 이 중에 더 정확하게 납치상황에 해당하는 프레임임을 인식하는 모델을 선정하여 납치상황 인식을 위한 프레임별 분류모델로서 사용하였다.

그 결과 Table 4에 보이는 바와 같이 J48 decision tree가 가장 높은 성능을 보였으며, TPR(True Positive Rate)이나 FPR(False Positive Rate) 측면에서도 다른 분류기법을 기반으로 한 모델보다 좋은 성능을 보였다. 이에 따라 J48을 사용한 납치 인식 모델을 본 연구의 납치 상황 인식을 위한 분류 모델로서 활용한다. 이때 평가 지표에 활용한 정확도(Accuracy), TPR 그리고 FPR은 아래 식 (4), (5), (6)과 같이 구한다[17].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN} \quad (5)$$

$$False\ Positive\ Rate = \frac{FP}{FP + FN} \quad (6)$$

Table 4. Performance comparison result of the proposed kidnapping detection model

	J48	NaiveBayes	SVM
Accuracy	88.7%	78.1%	82%
TPR	86.0%	68.5%	80.8%
FPR	9.3%	10.8%	17.2%

IV. Experiment and evaluation

본 장에서는 앞서 제안한 납치 상황 인식 모델의 성능을 평가하기 위한 실험 및 결과 분석에 관하여 기술한다.

4.1 Data acquisition

실험에 사용할 데이터 취득을 위해 아래 Table 5의 카메라 1대를 인하대학교 캠퍼스내에 설치하였으며, 이를 이용해 납치 상황의 영상 85개, 일반 상황의 영상 84개, 총 169개의 실험영상을 취득하여 사용하였다.

Table 5. Camera used for data acquisition

Camera Model	Camera Type	FPS
TRUEN TN-P1222CPLX	PTZ	24

실험을 위한 납치 상황의 영상을 취득하기 위해서 YouTube

로부터 실제 납치 상황의 영상들을 검색하고 이를 참조하여 85건의 영상을 자체 제작하였으며, 또한, 일반 상황의 영상도 YouTube의 2사람 이상의 보행자가 동행하는 영상들을 참조하여 84건의 영상을 자체 제작하였다.

4.2 Performance evaluation

본 절에서는 앞서 선택한 J48 기반의 납치 상황 프레임 분류 모델을 바탕으로 납치 영상을 인식하는 납치 인식 모델의 성능을 평가하고 그 결과를 제시한다.

특정 영상이 납치인지 아닌지를 판단하는 방법은 다음과 같다. 특정 영상에 존재하는 모든 프레임에 대하여 아래 식 (7)과 같이 프레임 기반 납치 상황 인식 모델이 납치 상황에 해당하는 프레임으로 판단한 프레임의 수(K)를 전체 프레임의 수(N)로 나눈 비율 $(\frac{K}{N}) \times 100$ 이 특정 문턱치(Threshold) 이상인 경우 해당 영상을 납치 영상인 것으로 판단하고 그렇지 않은 경우 해당 영상을 납치에 해당하지 않는 것으로 판단한다. 이때의 문턱치는 실험을 통해 최적의 성능을 보이는 63으로 결정하였다.

$$(\frac{K}{N}) \times 100 \geq Threshold \quad (7)$$

이에 따라 앞서 학습한 프레임 단위의 납치상황 인식 모델과 문턱치를 바탕으로 영상 내 납치상황의 존재여부를 판단하는 실험을 진행하였다. 그 결과 Table 6에 나타난 바와 같이 정확도(accuracy) 97.63%, TPR 100%, FPR 4.5%의 성능을 나타내었으며, 이는 과거 제안한 기존의 납치 상황 인식 모델[4]보다 정확도 측면에서는 약 4.73%, TPR의 경우에는 약 6.1% 가량 성능이 증진된 결과이며 FPR의 경우에는 약 3.6%가량의 성능이 개선되었음을 확인하였다.

Table 6. Performance comparison between the proposed kidnapping detection scheme using human pose estimation and that of [4]

	Kidnapping detection scheme in [4]	New kidnapping detection scheme using human pose estimation
Accuracy	92.90%	97.63%
TPR	93.9%	100%
FPR	8.1%	4.5%

IV. Conclusions

본 논문에서는 지능형 감시 시스템에서 정확한 납치 상황 인식을 위해 사람 자세 추정 특징을 추가로 이용하는 납치 상황 인식기를 제안하였다. 기존 연구에서 사용한 납치범과 피해자의 행위 크기 및 규칙성을 표현하는 특징 정보 이외로 사람의

자세 추정 특징을 구하여 이용하기 위해서 입력 영상으로부터 OpenPose 라이브러리를 이용하여 사람 관절의 위치 정보를 추출하여 관절간의 거리, 속도, 폐색 여부 등을 표현하는 특징 정보를 계산하여 활용하였다. 사람의 자세 추정을 위하여 팔과 몸의 움직임 판단하기 위해 총 10개의 관절의 위치 정보를 구하여 유클리드 거리와 속도 그리고 폐색에 관한 21가지의 후보특징을 추출하고 특징 선택을 하여 납치 상황 인식에 도움이 될 만한 사람 자세 추정 특징을 4가지를 선정하여 기존의 납치 상황 인식 연구에서 행위 크기 및 규칙성 특징과 합쳐 총 8가지의 유효 특징으로 구성된 프레임 단위 납치 상황 인식 모델을 만들었다. 성능평가 결과에 따르면 본 논문에서 제안한 사람 자세 추정을 이용한 납치 상황 인식기가 기존의 납치 상황 인식기보다 4.73% 높은 정확도를 갖는 것으로 나타났다. 따라서 제안한 납치 상황 인식기는 사람의 관절 위치 정보를 납치 상황 인식기에 적용하여 납치 상황 및 일반 동행 상황에서 두 사람의 만남이 인식되는 순간부터 그들이 보이는 행위를 충분히 판단 가능하게 하였다.

향후 연구로는 인식 가능한 납치 상황의 다양성을 증진하기 위해서 납치범과 피해자의 수를 제안하지 않고 여러 납치범들이 등장하는 상황 또는 차량을 이용하는 납치 상황들을 인식할 수 있도록 개선할 예정이다. 또한, 납치 인식의 정확성 증진을 위해서 심층 학습(deep learning) 기법을 이용하는 납치 상황 인식기 개발을 위한 연구를 진행 할 예정이다.

REFERENCES

- [1] Statistics Korea, http://stat.kosis.kr/statHtml_host/statHtml.do?orgId=132&tblId=DT_13204_2011_211&conn_path=I2&dbUser=NSL_IN_132
- [2] Choong-Sik. Chung, "A Case Study on the Operation Enhancement of Integrated CCTV Control Center at Busan Metropolitan City," Journal of Korean Association for Regional Information Society, Vol 18, No. 3, pp. 123-154, September 2015.
- [3] YTN News: Detect immediately dangerous situations ... prevent crime with intelligent cctv, http://www.ytn.co.kr/_ln/0115_201212241855130011
- [4] Ryu-Hyeok. Gwon, et al, "A Kidnapping Detection Scheme Using Frame-Based Classification for Intelligent Video Surveillance," Proc. Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, pp. 345-354, October 2013.
- [5] Ji-Hyen. Choi, et al, "A Prediction Method for Abnormal Behavior based on Omni-view Pattern," Proc. of the 42th KIISE Winter Conference, pp. 401-403, Korea, 2015.
- [6] Soumi. Paul, et al, "Microsoft Kinect in Gesture Recognition: A Short Review," International Journal of Control Theory and Applications, Vol. 8, No. 5. pp. 2071-2076, 2015.
- [7] Shih-En. Wei, et al, "Convolutional pose machines," IEEE Conference on Computer Vision and Pattern Recognition, pp. 4724-4732, 2016.
- [8] Zhe. Cao, et al, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields" IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291-7299, 2017.
- [9] OpenPose library, <https://github.com/CMU-Perceptual-Computing-Lab/openpose>
- [10] Divya. J, "Automatic Video Based Surveillance System for Abnormal Behavior Detection," International Journal of Science and Research, Vol 4. No 7, pp 1743-1747, July 2013.
- [11] Hall. Mark A, et al, "The WEKA data mining software: An Update ACM SIGKDD explorations newsletter," Vol. 11, No. 1, pp. 10-18, June. 2009.
- [12] Hall. Mark A, "Correlation-based feature selection for machine learning," Ph. D. Thesis, University of Waikato, pp. 1-198, Apr. 1999.
- [13] Lei Xu, et al, "Best first strategy for feature selection," 9th International Conference on Pattern Recognition, pp. 706-708, Italy, 1988.
- [14] Kevin P. Murphy, "Naive Bayes classifiers," University of British Columbia, Vol. 18, Oct. 2006.
- [15] Neeraj. Bhargava, et al, "Decision Tree Analysis on J48 Algorithm for Data Mining," Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3. No. 6, pp. 1114-1119, 2013.
- [16] Marti A. Hearst, et. al, "Support vector machines," IEEE Intelligent Systems and their applications, Vol. 13, Issue. 4, pp. 18-28, 1998.
- [17] Powers, et. al, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," Journal of Machine Learning Technologies, Vol. 2, No. 1, pp. 37-63, February 2011.

Authors



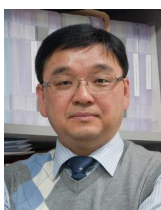
Ju Hyun Park received the B.S. degrees in Computer Science Engineering from Incheon University, Korea, in 2015. Mr. Park is currently a M.S. student in the Department of Information and Communication Engineering, Inha University. He is

interested in data mining, computer vision, and big data.



KwangHo Song received the B.S. and M.S. degrees in the Information and Communication Engineering from Inha University, Korea in 2015 and 2017, respectively. Mr. Song is currently a Ph.D. candidate in the Department of Information

and Communication Engineering, Inha University. He is interested in data mining, text mining, and big data.



Yoo-Sung Kim received the B.S. degree in Computer Science from Inha University, Korea, in 1986. He received M.S. and Ph.D. degrees in Computer Science from KAIST, Korea, in 1988 and 1992, respectively. Dr. Kim is currently a professor in the

Department of Information and Communication Engineering, Inha University. He is interested in multimedia information mining, big data, and intelligent video surveillance system.